VALUE IN HEALTH 16 (2013) 259–266

Available online at www.sciencedirect.com

SciVerse ScienceDirect

journal homepage: www.elsevier.com/locate/jval

Evaluating the Impact of Unmeasured Confounding with Internal Validation Data: An Example Cost Evaluation in Type 2 Diabetes

Douglas Faries, PhD[1,], Xiaomei Peng, MS[1], Manjiri Pawaskar, PhD[1], Karen Price, PhD[1], James D. Stamey, PhD[2], John W. Seaman Jr., PhD[2]*

[1]*Eli Lilly & Company, Indianapolis, IN, USA;* [2]*Department of Statistics, Baylor University, Waco, TX, USA*

ABSTRACT

The quantitative assessment of the potential influence of unmeasured confounders in the analysis of observational data is rare, despite reliance on the "no unmeasured confounders" assumption. In a recent comparison of costs of care between two treatments for type 2 diabetes using a health care claims database, propensity score matching was implemented to adjust for selection bias though it was noted that information on baseline glycemic control was not available for the propensity model. Using data from a linked laboratory file, data on this potential "unmeasured confounder" were obtained for a small subset of the original sample. By using this information, we demonstrate how Bayesian modeling, propensity score calibration, and multiple imputation can utilize this additional information to perform sensitivity analyses to quantitatively assess the potential impact of unmeasured confounding. Bayesian regression models were developed to utilize the internal validation data as informative prior distributions for all parameters, retaining information on the correlation between the confounder and other covariates. While assumptions supporting the use of propensity score calibration were not met in this sample, the use of Bayesian modeling and multiple imputation provided consistent results, suggesting that the lack of data on the unmeasured confounder did not have a strong impact on the original analysis, due to the lack of strong correlation between the confounder and the cost outcome variable. Bayesian modeling with informative priors and multiple imputation may be useful tools for unmeasured confounding sensitivity analysis in these situations. Further research to understand the operating characteristics of these methods in a variety of situations, however, remains.

Keywords: Bayesian methods, confounding, robustness, sensitivity analyses.

Introduction

The use of retrospective observational research as a tool for medical decision making, particularly with data from health care claims databases and electronic medical records, has been growing in recent years. With large and heterogeneous populations of patients, such observational databases are a rich source of usual care data, which can potentially address a variety of medical questions [1,2]. The use of such data for comparative effectiveness, however, is challenged by selection bias and potential for unmeasured confounding [3–5]. Patients are not randomized to treatments and thus comparisons between treatment groups are subject to bias due to the many factors that influence treatment choices in usual care practice. Statistical adjustment for measured confounders is possible, such as through propensity score adjustment. The validity of such methods, however, relies on the assumption that there are no unmeasured confounders. That is, there are no factors related to both treatment and outcome that are not collected and appropriately utilized in the analysis. As this assumption cannot be verified, observational data have lower internal validity and are lower on the hierarchy of evidence relative to randomized clinical trials [6–9].

In prospective observational studies, a researcher can specify the collection of data on known confounders; however, this opportunity does not exist in retrospective database research. While researchers look for proxies for such known confounders within the existing database, the degree to which this addresses the confounding is unknown. In addition, unknown confounders may exist and without randomization such variables will cause the standard analyses to be biased. To ensure the robustness of the observational research findings, it is important to conduct sensitivity analyses to assess the potential impact of unmeasured confounding [4,9–10].

While many researchers mention the limitations on inferences from their work due to unmeasured confounding, few directly assess the potential impact in a quantitative fashion [10,11]. Even when no or limited additional data on the unmeasured confounders are available, there are several methods that can be utilized to assess sensitivity for unmeasured confounding, including the Rule Out [10] and Bayesian modeling with noninformative priors [12]. The Rule Out approach uses a simple model to quantify the level of unmeasured confounding necessary to eliminate the observed treatment difference (e.g., moves the risk ratio to 1). Researchers can then assess whether such a level of confounding is plausible for their scenario.

*Address correspondence to: Douglas Faries, Eli Lilly & Company, Lilly Corporate Center, Indianapolis, IN 46285, USA
E-mail: d.faries@lilly.com.

1098-3015/$36.00 – see front matter

http://dx.doi.org/10.1016/j.jval.2012.10.012

Data on a known unmeasured confounder may be available external to the study ("external information"), such as in a related study or survey data [13–15]. Bayesian approaches [11,12,16] are a natural fit as the uncertainty in the "external information" can be modeled as part of an informative prior distribution. The use of external data, however, requires transportability between populations and typically faces difficulty accounting for the correlation between unmeasured confounders and measured covariates already in the original analysis. The impact of an unmeasured confounder can clearly depend on the extent to which a missing variable is related to other variables in the analysis. This problem can be addressed directly when "internal information" is available. For example, chart review data could be obtained on a subsample of the patients in administrative claims database analysis and could be used as "internal information."

Methods utilizing internal data include Bayesian modeling, propensity score calibration (PSC) [17], and multiple imputation [18,19]. Bayesian approaches can utilize the internal data in their prior structures. PSC is based on measurement error methodology and accounts for the correlation between the confounder and other variables already in the full model. The method, however, depends on the assumption that the association between the unmeasured confounder and outcome is in the same direction as the association between the measured confounders and the outcome. This clearly will not be valid in all scenarios, and the method performs poorly when this assumption is violated [20].

When some internal data on the "unmeasured" confounder exist, the issue can be viewed as a missing data problem, and methods such as multiple imputation can be utilized. Incorporating the outcome variable in the imputer's model allows for full flexibility in regard to the relationship of the unmeasured confounder with the outcome variable—a relationship that is restricted in the PSC approach. The properties of such multiple imputation approaches when the majority of the data is missing, as in "unmeasured" confounding scenarios, however, are not well known. This multiple imputation approach also falls within the general probabilistic resampling framework discussed by Gustafson and McCandless [11].

A recent study [21] compared health care costs for patients initiating exenatide (referred to as Treatment 1 throughout the remainder of this article) versus insulin glargine (Treatment 2). This was a retrospective claims database study, and information on some potential confounders, such as the level of glycemic control (measured by glycoslylated hemoglobin [Hb], or Hb A_{1c}) was minimal. Indeed, Hb A_{1c} values were recorded only for 25% of the patients during the preinitiation period, and so there was no adjustment for glycemic control in the primary analysis. In this study, we performed a post-hoc analysis to demonstrate methods for assessing the impact of the unmeasured confounding in the case in which there is internal data—such as glycemic control in this example. Specifically, three different methods incorporating the internal data were assessed: 1) Bayesian modeling, 2) multiple imputation, and 3) PSC. A Rule Out approach was also used as an initial assessment of the robustness of the results to potential unmeasured confounding.

Methods

Database

The objective of the original analysis was to compare the total health care costs for type 2 diabetes patients initiating one of two treatments with a 12-month follow-up period. The analysis was performed by using an administrative claims database from i3 InVision. This database includes medical and prescription claims from more than 30 million patients throughout the United States during the study time period. See Pawaskar et al. [21] for further details concerning the database and population for this analysis. In brief, patients were included in this study if they received at least one new prescription for either of the diabetic medications between April 1, 2005 (the earliest month both medications were on the market in the United States), and June 30, 2007 (the latest available data at the time of the original analysis), and must have had at least one diagnosis for type 2 diabetes identified by using *International Classification of Disease, Ninth Revision, Clinical Modification* during the preindex period. The first prescription date for study medication was identified as the index date. The preindex period was defined as a 6-month period prior to the index date (i.e., 6 months before the first prescription of index medication) and the postindex period was defined as a 12-month follow-up period after the index date. The database contained 93,345 patients initiating either treatment during the time period, with 10,074 meeting all inclusion/exclusion criteria (the largest exclusion of more than 41,000 patients due to a diagnosis of type I or gestational diabetes). Of the 10,074 patients in the analysis data set, 7,255 patients were in the Treatment 1 cohort, 2,819 in the Treatment 2 cohort, and baseline Hb A_{1c} data were available for only approximately 25% of these patients.

Adjustment and Outcomes Variables

Total health care cost over the 12-month follow-up period was the primary outcome variable for the previous study and thus is the outcome of interest for this analysis. The assessment of cost data adds additional analytic challenges in part due to the skewness of the data [22,23] but is an important outcome for the health care payer. Patients were not randomly assigned to treatment groups. Consequently, many differences between the populations of patients prescribed each treatment in usual care were expected and adjustment for baseline differences was necessary. The health care claims database allowed for the assessment of patient demographics, comorbidities, complications, resource use, and costs of care in the 6-month preinitation period. No data on the other potential confounders such as weight, body mass index (BMI), and duration of diabetes, however, were available. Only limited data on glycemic control, measured by Hb A_{1c}, were available in the subset of patients.

Statistical Methods

In the original analysis, the two treatment cohorts were matched by using a propensity score greedy 1:1 algorithm. The differences in total costs in the 12-month follow-up period between treatments cohorts were estimated by using a nonparametric bootstrapping test. The propensity model included patient demographics, general health status (measured by Charlson comorbidity index), medical comorbidities, diabetes-related complications, medication use (including prior antidiabetic medications), health care resource utilization (hospitalizations, emergency room visits, endocrinologist visits), and medical costs during the 6-month preindex period, but not Hb A_{1c}. Sensitivity analysis including propensity score stratified bootstrapping [24] and generalized linear regression models [25] was used to ensure the generalizability of results from matched cohorts to the entire study population. This study also performed a separate generalized linear model in the subgroup of patients for whom a preindex Hb A_{1c} value was available, to estimate the mean total costs when controlling for patients' glycemic control at the baseline.

The current analysis considered several other techniques to assess the potential impact of the exclusion of glycemic control information from the original analysis: 1) Bayesian modeling with internal validation, 2) multiple imputation, and 3) PSC. As the first step of the analysis, a Rule Out approach was included.